



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2015

Becoming aware of one's thoughts : Kant on self-knowledge and reflective experience

Renz, Ursula

DOI: <https://doi.org/10.1515/9783110378795.581>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-136732>

Book Section

Originally published at:

Renz, Ursula (2015). Becoming aware of one's thoughts : Kant on self-knowledge and reflective experience. In: Moyal-Sharrock, Danièle; Munz, Volker; Coliva, Annalisa. Mind, Language and Action. New York: De Gruyter, 581-599.

DOI: <https://doi.org/10.1515/9783110378795.581>

Ursula Renz

Becoming aware of one's thoughts

Kant on self-knowledge and reflective experience

Ursula Renz: University of Klagenfurt, ursula.renz@aau.at

In the last fifteen years, several authors have suggested we conceive of self-knowledge in terms of self-constitution. In developing this idea, most accounts rely on two presuppositions, one concerning the relation between the knowing subject and the mental states in question, and one concerning the importance of self-knowledge for human agency. In his *Authority and Estrangement*, for instance, Richard Moran famously contends that knowledge of one's own propositional attitudes is not merely an empirical matter, but involves a constitutive aspect. This, however, neither supports „a conventionalist reconstruction of first-person authority“ nor amounts to „a deflationary analysis of the claims of self-knowledge“ (Moran 2001, 26). Instead, Moran's view preserves a robust notion of first-person authority, and provides us with a preliminary explanation of how the capacity for self-consciousness first makes it possible for a person to be responsible for his mental life and actions.

Since the appearance of *Authority and Estrangement*, this view has been sharply criticized. Recently, however, it has been defended by Matthew Boyle, who points out some parallels between Moran's and Kant's approaches. Boyle claims that many problems raised in discussions about self-knowledge can be solved if one accepts a distinction between two kinds of self-knowledge similar to the Kantian opposition between self-knowledge through pure apperception and self-knowledge through inner sense. Boyle thus rejects what he calls „the Uniformity Assumption“; this is „the demand“ underlying many approaches „that a satisfactory account of our self-knowledge should be fundamentally uniform, explaining all cases of 'first-person authority' in the same basic way“ (Boyle 2009, 141). And he suggests that in contrast to most theorists, Moran is not bound to the Uniformity Assumption, although he does consider our deliberative knowledge of our own attitudes to be fundamental.

I am quite sympathetic to Boyle's intention. It is indeed a widespread tendency in philosophical discussions about self-knowledge to assume that the riddle of the (seeming or real) epistemic authority of the first person with respect to knowledge of one's own mental states can be solved by appealing to

one simple human capacity. And I agree with Boyle that this is a problematic presupposition. Furthermore, I share the view put forth by Boyle and Moran, as well as many others, that (roughly) Kantian approach may help us get a better understanding of the relation between self-knowledge or self-consciousness and self-constitution.¹

However, I am afraid that the way Kant is referred to in the defense of self-constitutionalism may hide another important aspect of Kant's views on self-knowledge. There is a way in which, according to Kant, people *passively* become aware of their mental life, *including* the thoughts they entertain. This is probably irrelevant for Kant's primary epistemological project, which aims to clarify the conditions of the possibility of knowledge. But it is nonetheless an important ingredient in Kant's overall outlook, insofar as it is a necessary condition for the possibility of enlightenment. That there is such a passive way of becoming aware of one's thoughts, is, I will argue, crucial for the understanding of how self-reflection may amount to a correction of one's attitudes.

In this paper, I want to substantiate this concern by providing an alternative reconstruction of Kant's view on self-knowledge. In highlighting this passive aspect of self-knowledge in Kant's approach, I want to draw attention to an issue often neglected in philosophical debates about self-knowledge. If we regard enlightenment as a real option for humans, then a philosophical model of human self-knowledge has to satisfy not just one, but two requirements. It must account for the first-personal nature of self-knowledge and it must do so in a way that makes it comprehensible how subjects may correct their epistemic or moral attitudes.

I proceed in several steps. In sections one to three, I try to make room for the view that Kant allows for the notion of our becoming aware of our mental properties, including our thoughts, in a passive way. Starting with some preliminary remarks on the textual evidence, I draw attention to two sets of interpretive questions that every philosopher has to answer when proposing a Kantian style theory of self-knowledge. Sections two and three discuss Boyle's answers to these questions and point out alternative ways of answering these two sets of question. It is against this background that I address my main interpretive ques-

¹ It has to be emphasized that reference to a Kantian idea of self-constitution was used by many to overcome the rather conventionalist view of self-constitution featured in the "Wittgensteinian legacy" in Wright (1998). Besides Moran (2001), Bilgrami (2006) and Rödl (2007) offer versions of Kantian-style constitutionalist views on self-knowledge. It is, however, Boyle (2009) who most explicitly and carefully exposes what could be considered the "Kantian legacy" concerning the issue of self-constitution. The relation between self-consciousness and self-determination in Kant is also pointed out by Allison (2006).

tion in section four. How is it possible, according to Kant, that we sometimes come to know our thoughts in a passive way, and why does this matter? As can be shown by the famous reflection R 5655 on *Whether it is an experience that we think*, Kant does assume that we sometimes perceive a thought in a passive way, or through inner sense. I argue that this is not only consistent with the reading of Kant's view on self-knowledge spelt out in section two and three, but it is, moreover, crucial to what one could call Kant's second epistemological project, namely that of making processes of enlightenment conceivable. I end in section five by arguing that this perspective should be considered crucial for a philosophical theory of self-knowledge.

1 Two sets of interpretive questions

As is well known, Kant did not present his views on self-knowledge in a separate treatise. To reconstruct his account of self-knowledge, one has to rely on several remarks scattered among different texts, both published and unpublished. Unfortunately, these remarks do not provide a unified picture, but they suggest that Kant's writings on self-knowledge can be divided in the same way his overall philosophy can be: into a pre-critical, a critical, and a late period.² In what follows, I rely primarily on the views of the second period.

There, Kant is mostly concerned with explicating and defending his distinction between two basic kinds of self-consciousness, named 'pure apperception' and 'inner sense'.³ This emphasis is not surprising, for the distinction underlies

² The early views, which Kant held until as late as 1778, are contained in passages from the *Lectures on Metaphysics* and the *Duisburger Nachlass*; see Kitcher 2011, 66-77, for an analysis of them. Important texts for the critical period are the chapters on the Transcendental Deduction and the Paralogisms of pure Reason in the second edition of the *Critique of Pure Reason*, § 46 of the *Prolegomena*, and §§ 1-7 and 24 in the *Anthropology from a Pragmatic Point of View*. The late period is represented primarily by the passages on self-positing in the *Opus postumum* (primarily fascicle VII and X). The posthumous manuscript *On Inner Sense* can be considered transitional; it introduces a shift between the second and the third periods. See Klemme (1996, 5-17) and Zobrist (2011) for an analysis of the development of Kant's views on the subject.

³ Unlike in contemporary debate, where 'self-knowledge' is often used as co-extensional with or even equivalent to 'self-consciousness' or 'self-awareness', Kant draws a clear distinction between the two concepts such that self-consciousness, or rather the two kinds of self-consciousness, is only a precondition of what he calls self-knowledge. The latter is distinguished from mere self-consciousness by the requirement that it has to satisfy Kant's rather demanding conditions for a thought to count as knowledge. These differentiations are im-

the arguments of some of most central passages of the *Critique of Pure Reason*. One could even say that, to some degree, it was his insight into the consequences of this distinction that forced Kant to revise the chapters on the Transcendental Deduction and the Paralogisms of Pure Reason in the second edition of the First Critique.⁴

How is this distinction to be understood? To begin with, let us have a closer look at the passages in which the two conceptions of self-awareness are introduced. Most crucial for the notion of transcendental apperception is the following famous passage in § 16 of the B-Deduction, which says

The I think must be able to accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least would be nothing for me.⁵

The concept of inner sense, on the other hand, is introduced already in the Transcendental Aesthetics, where Kant says,

Inner sense, by means of which the mind intuitively itself, or its inner state, gives, to be sure, no intuition of the soul itself, as an object; yet it is still a determinate form, under which the intuition of its inner state is alone possible, so that everything that belongs to the inner determination is represented in relations of time.⁶

The distinction between the two kinds of self-awareness, finally, is extensively discussed in both the Transcendental Deduction and the Paralogisms of Pure Reason, where Kant emphasizes several times that it is only in virtue of the transcendental apperception that we can think of ourselves as the identical subject of our mental properties, whereas inner sense provides us only with an empiri-

portant for the point I am going to make; I am thus following Kant's terminology rather strictly henceforth, though using 'self-consciousness' and 'self-awareness' interchangeably.

⁴ See Horstmann 1993, Brook 2013, and Brandt 2013 for analysis and discussion of the differences between the two editions related to the issue of self-consciousness.

⁵ „Das: Ich denke, muß alle meine Vorstellungen begleiten können; denn sonst würde etwas in mir vorgestellt werden, was gar nicht gedacht werden könnte, welches ebensoviele heißt, als die Vorstellung würde entweder unmöglich, oder wenigstens für mich nichts sein.“ (B 131f., the English translation quoted in the main text is from the edition by Paul Guyer and Allen W. Wood.)

⁶ „Der innere Sinn, vermittels dessen das Gemüt sich selbst, oder seinen inneren Zustand anschaut, gibt zwar keine Anschauung von der Seele selbst, als einem Objekt; allein es ist doch eine bestimmte Form, unter der die Anschauung ihres inneren Zustandes allein möglich ist, so daß alles, was zu den inneren Bestimmungen gehört, in Verhältnissen der Zeit vorgestellt wird.“ (B 37)

cal consciousness of ourselves insofar as we appear to ourselves. It is thus impossible for humans to have a notion of the I intuitively or, what amounts to the same idea, to perceive themselves as the subjects of their mental properties through inner sense. The overall picture in Kant's critical philosophy is thus this: Kant assumes that, corresponding with the two sources of knowledge, there is a categorical distinction between two kinds of self-consciousness, each of which has a different function in both our epistemological and moral self-reflection.

While the general outlook seems clear, there are several issues with regard to which Kant's approach is in need of further clarification. These may be divided into two sets of questions that one has to answer when interpreting Kant's views on self-knowledge. First, one might wonder how to explicate the two notions of self-consciousness in detail. There is thus a first kind of interpretive concern exhibited by the following questions:

(1) What kind of entity (or kinds of entities) does Kant refer to when he distinguishes the kind of self-consciousness contained in pure apperception from the kind provided through inner sense? Do these notions denote two distinct *modes of representation* involved in self-reflection only, or do they differentiate between two types of *objects* of human self-consciousness? Are there, moreover, strict correlations between modes of self-reflective representation and types of objects? Clearly, given the brief exposition above, there is one strict *negative* correlation: inner sense cannot possibly represent the I, the subject of mental properties as it is in itself. But what about the correlations between modes of self-reflective representation and different types of mental properties? Is there, for instance, a strict relationship between different modes of representation and types of mental properties such that it is only in virtue of pure apperception that one's thoughts may be represented in self-reflection? Or is there room for contingency with respect to how certain kinds of mental processes are represented in self-reflection? How, finally, is Kant's emphasis on the difference between the *activity* of pure apperception and the experiential *passivity* through which we acquire self-knowledge in inner-sense to be understood?

Note that all these questions are related to definitional issues. They arise because they are either not explicitly addressed by Kant himself, or because they are dealt with in a less than clear manner. But one cannot leave them unanswered. Depending on how precisely they are answered, one may also question whether or not there are only two kinds of self-reflective representations in Kant. I will come to this issue in the next section. First, however, I would like to point to another set of interpretive questions, which are related to Kant's concept of knowledge. Given the different functions of pure apperception and inner sense, one can assume that there are some dependencies between the two kinds

of self-consciousness. Most famous is the suggestion in the sentence quoted above, that purely apperceptive self-consciousness is a necessary condition for all kinds of knowledge, including, of course, the kind of empirical self-knowledge that consists in the categorization and determination of one's own mental states.

We may thus specify a second set of interpretive questions, which is concerned with the relations between the two kinds of self-consciousness, pure apperception and inner-sense:

(2) What kind of dependency is there between the two kinds of self-consciousness, and how is it argued for? Is the primacy which Kant attributes to pure apperception in the Transcendental Deduction an expression of the epistemological program pursued in the *First Critique*, or is it a requirement of his conception of the human mind? Is it, in other words, due to Kant's transcendental idealism, or is it a consequence of Kant's view of the mind as having different ontological layers such that the existence of self-consciousness through inner sense is conditioned by the existence of purely apperceptive self-consciousness? If the latter, what exactly is required by this claim? What does "existence" mean in this context? Must a subject, in order to have some inner-sense experience, *actually perform* an act of pure apperception, or is it sufficient that she have *the capacity* for this kind of self-consciousness?⁷ Finally, is this primacy the only relation Kant assumes between these two kinds of self-consciousness?

It is not obvious how Kant would have answered all these questions. Both sets of questions have aroused exegetic controversies.⁸ Since I am concerned with the issue of self-knowledge, I am primarily interested in the theoretical options related to these questions. The controversies surrounding Kant's actual views on these issues show, I think, that there are more possibilities for theoretical differentiation than is often assumed. These might turn out to be relevant for the debate about self-knowledge as well. I will discuss this at greater length in the rest of this paper. At this point, I only want to mention my major concern:

⁷ I follow here Emundts (2013), who has recently distinguished between two readings of Kant's notion of self-knowledge through pure apperception, i.e. a reading according to which Kant is merely referring to our faculty of thinking, and a reading which takes self-knowledge through pure apperception to consist in some kind of actual consciousness. Albeit crucial, this distinction is only rarely explicitly discussed. As I am going to point out, this is also the case in Boyle's defense of Moran. In contrast to Emundts, however, I am not primarily interested in providing a true interpretation of Kant's views. Instead, I make use of her distinction in order to point out different types of broadly Kantian theories of self-knowledge.

⁸ See Emundts (2013b, 52f.) for controversies surrounding the first set of questions, Horstmann (2013) for those surrounding the second set of questions.

I am worried that in recent debates on self-knowledge, the focus on self-constitution leads to the subject's activity in achieving self-knowledge being overemphasized, while those aspects of self-consciousness and self-reflection that are irreducibly passive are marginalized or even misunderstood. To see how Kant deals with the above questions may help to understand why and in what respect passivity matters.

2 Varieties of self-representation

As already mentioned, Boyle's primary goal in referencing Kant is not to engage in an exegesis of Kant's arguments, but to put Moran's notion of deliberative self-knowledge into a broader framework. In doing so, Boyle provides a new interpretation of the Kantian dichotomy between pure apperception and inner sense, according to which this dichotomy is introduced to distinguish between two forms of self-knowledge, an active and a passive form of self-knowledge. Kant's intuition behind this move consists, Boyle suggests, in the assumption that only a mind capable of active self-knowledge, of thinking of oneself as the maker of one's thought, is also capable of passively receiving some representation of its own occurrent mental states in inner sense. In emphasizing this dependency between the two kinds of self-knowledge, Boyle takes Kant's affirmation of the priority of pure apperception to be constitutive of what may be called a mature human mind. Being self-knowers, mature human minds are subjects who have an understanding of the concepts involved in their judgments or attitudes, and this requires that they are able to form judgments or adopt attitudes on the basis of rational deliberation. On the other hand, subjects lacking such an understanding of concepts cannot even legitimately be considered knowers.⁹

How does this interpretation respond to the questions raised in the previous section? As to the first set of questions, Boyle's view seems to be that the distinction between pure apperception and inner sense sets apart both modes of presentation and types of objects of human self-knowledge. Correspondingly,

⁹ Following this picture of Kant, the crucial point of Moran's notion of self-constitution lies in the conception of the human or rational mind provided thereby and not in the description he gives of the process of self-knowledge. It is in virtue of being able to take a deliberative stance towards one's own attitudes that humans are rational subjects. This is surely an adequate and charitable picture of Moran's contention, as can be shown by the fact that Moran himself appeals to Kant more than once in order to clarify the point of his distinction between the stance of causal explanation and the stance of rational agency; cf. Moran (2001, 89 and 127).

he assumes that there is a strict correlation between modes of self-representation and objects, such that mental activities are the objects of the active mode of representation, whereas passively acquired mental states have to be represented passively. He thus ascribes to Kant the view that we know our thoughts exclusively through pure apperception, whereas our sensations and emotions are known through inner sense.

Relying on this view, we may also explicate Boyle's answers to the second set of questions. In accordance with the assumption of a strict correlation between modes of presentation and objects of self-reflection, Boyle seems to suggest that the primacy of pure apperception as maintained in the Transcendental Deduction is a matter of the constitution of mature human minds. For a subject to be a mature human mind, it is necessary that it have the capacity for pure apperception or, as the parallel with Moran's approach suggests, for rational deliberation of one's attitudes. This also suggests that man's capacity for inner sense is shaped by his capacity for pure apperception without, however, being reducible to it. Boyle thus seems to read Kant's claim of the primacy of pure apperception as an anthropological tenet based on a dependency relation between two kinds of capacities rather than actual exercises of these capacities.¹⁰

Boyle's reconstruction of Kant is very clear, and it is characterized moreover by articulate commitments. He explicitly argues that the decisive contrast between the two kinds of self-knowledge is a matter of capacities and not of occurrent uses of the capacities, and he obviously adopts an anthropological rather than an epistemological reading of Kant. One might wonder, however, whether this does not hide some of the complexities of Kant's views. There is in particular one point at which Boyle's exposition departs substantially from Kant's own articulation of his account. Explaining the distinction between pure apperception and inner sense, Boyle draws on some formulations of Kant appearing in § 24 of the *Anthropology*. There Kant says:

Der innere Sinn ist nicht die reine Apperception, ein Bewußtsein dessen, was der Mensch thut, denn dieses gehört zum Denkungsvermögen, sondern was er leidet, *wiefern* er durch sein eignes Gedankenspiel afficirt wird (AA 7: 161).

Boyle renders this passage as follows:

10 The idea that this dependency relation is essential to the nature of human subjects is also maintained by Kitcher. In contrast to Boyle, however, she emphasizes that this dependency is not merely a matter of capacities, but of activities: "Absent the activities involved in judging and inferring, humans would remain in a state of being potential thinking selves; it is only the exercise of these capacities that creates the relations among mental states that are distinctive of cognitive subjects and that permit them to understand themselves as 'I's'" (Kitcher 2011, 264f.).

Kant says: that whereas our apperceptive knowledge of our own judgments is a knowledge of 'what we do [thun]', our knowledge of our sensations and appetites through inner sense is a knowledge of what we 'undergo [leiden].'

In the next sentence he recalls this passage by referring to the "Kantian contrast between an active and a passive form of self-knowledge" (Boyle 2009, 158).

This is not precisely what Kant says. First, Kant is not explaining the notion of inner sense in terms of our awareness of sensations and appetites; he doesn't even mention our sensations and appetites in this paragraph. Instead, he equates pure apperception with man's consciousness of what he does, and he contrasts this with his awareness of what he undergoes insofar as he is affected by the play of his own thought. Note that Kant uses the terminology of thought in both formulations: pure apperception belongs to the capacity of thought, but what we undergo is also an instance of thought, or of the play of our own thought. Taking this seriously, we have to conclude that Kant's concept of thought allows for two options that are missed in Boyle's paraphrase of the quoted text. First, we can be affected by our thought or by the results of our thinking activity; second, it must be possible for these to be known in a passive way.

There is another peculiarity in Boyle's exposition. In the text mentioned above, Kant indeed contrasts passivity and activity, but only at level of the *objects* of self-consciousness. This is not equivalent to the contrast "between an active and a passive *form* [emphasis U.R.] of self-knowledge", as Boyle puts it. This is not to say that Kant does not sometimes distinguish pure apperception and inner sense by characterizing the first as an active and the second as a passive form of self-consciousness. But so far as I can see, he is not strictly correlating consciousness of what we do with an active, and self-consciousness of what we undergo with a passive, form of self-reflection.

But, one might say, isn't this correlation involved in Kant's categorical denial that inner sense is able to represent the I or the subject of one's mental properties? I don't think so. Clearly, given this denial, I cannot endorse some instance of my activity *as mine* by perceiving it through inner sense. But this neither entails that I am unable to sense instances of thought as mental states I am undergoing, nor does it preclude that I am perceiving these through inner sense. On the contrary, it seems hard to make sense of the idea that we can be affected by the play of our thoughts and be aware of this by inner sense if these options were ruled out.

This indicates that Kant's approach contains more theoretical differentiations than one might think if one focuses simply on the distinction between an active and a passive form of self-consciousness. We have to presume that the

dichotomy between activity and passivity is applied to both the objects and the forms of reflective representation. This allows us to combine the types of mental capacities or processes constituting the object of self-reflection with the modes of representing them in self-reflection in several ways. This can best be illustrated by the following cross classification:

object	active	passive
mode of representation		
active	(1) taking a stance towards some thought	(2) taking a stance towards some feeling or sensation
passive	(3) perceiving my being affected by some thought ¹¹	(4) perceiving my being affected by some feeling or sensation

Following this classification, there are at least four options for how humans may self-consciously represent their own minds: (1) they can take a stance towards some thought; e.g. by actively affirming or denying it; (2) they can take a stance towards some sensation or emotion; e.g. by actively approve or condemn it; (3) they can perceive their being affected by some thought; e.g. by stating its occurrence; and (4) they can perceive their being affected by some sensation or feeling; e.g. by just noting it.

Note that these examples are only meant to illustrate the distinctions in question; they are mine and not Kant's. It has to be admitted furthermore that not all the options are of the same importance in Kant's transcendental philosophy. It is not clear, for instance, whether (2) plays any significant role in the *First Critique*, and if so, on which level. Also, one might wonder whether the claim that "[t]he I think must accompany all my representations" refers to (1) only or to (1) and (3) together. But this is not to say that Kant did not develop his views against the background of the full spectrum of possibilities as depicted by

¹¹ Yoon Choi has raised the question of whether the content of the perception includes "my being affected" by the thought or the sensation. I think it contains my "being affected," which is ipso facto "my being affected" but definitely not "me being affected".

this classification, on the contrary. I will argue in the last section that there are at least good reasons to assume that (3) plays an important role for Kant as well.

3 The primacy of pure apperception

As previously mentioned, Boyle is committed to a reading of Kant that assumes some categorical dependency of inner sense on pure apperception at the level of man's mental capacities. The argument for this assumption is provided by what Boyle describes as a minimal condition for self-knowledge. According to this requirement, every subject capable of knowing herself must have the capacity to represent her own condition as being of a certain kind (Boyle 2009, 143 and 161). Otherwise we (and she) would not be able to understand what makes the object of self-knowledge a self (Boyle 2009, 161). This minimal condition for self-knowledge parallels Moran's notion of authoritative self-knowledge: A subject is capable of representing her own condition if and only if she is able to take a deliberative stance towards her attitudes. At the same time, this minimal condition is used to explain Kant's claim that all our empirical consciousness contained in inner sense is conditioned by the consciousness of the I provided by pure apperception.

This does indeed capture an important point. Kant does assume that in pure apperception we do not simply contemplate our mental states and activities, but actively exercise our capacity to combine representations and thereby affirm our authority or, in Kant's terms, our spontaneity which we have in virtue of being the makers of our thoughts. Furthermore, the awareness of our being the makers of our thoughts is indeed what constitutes the essence of the self. "[T]he I", Kant says in the paralogisms, "is only the consciousness of my thinking" (B 413); here thereby denies any substantial self.

I nonetheless think that the point behind Kant's central claim in § 16 of the B-Deduction (B 136) that it must be possible for the I think to accompany all my representations is different than Boyle's anthropological interpretation suggests. There are two valid readings of this sentence, to my mind. One says that wherever my representations originate from, none of them can constitute an objective thought unless I can accompany it by the thought 'I think'. What is required, according to this interpretation, is not the capacity for pure apperception as such, but the capacity for accompanying the particular representations in question by some act of apperception. Given this first reading, Kant's concern in emphasizing the dependency of all our representations on pure apperception is not simply whether we have a certain cognitive ability, but whether a repre-

sensation is available to us in a sufficiently clear and distinct form such that it can be picked out in an act of thought.¹² Otherwise, I am not in the position to use it within the formation of some judgment. The second reading departs even further from Boyle's suggestion. It says that a representation received in inner sense cannot constitute an objective thought in the absence of that self-consciousness which is constituted by the activity of pure apperception. Given this interpretation, any instance of self-knowledge requires that we *actually have* the relevant self-consciousness. I do not want to decide here which of these readings is true. Yet according to both these readings, Kant's point is more specific than Boyle's anthropological reconstruction suggests. Both relate the sentence in §16 of the B-Deduction to the epistemological program of the *First Critique*, which aims to explicate the conditions of our knowledge of unified objects. Unlike Boyle's account, these readings do not ascribe to Kant a particular conception of the human mind.¹³

Given the epistemological program of the *First Critique*, both of these readings are, I think, preferable to Boyle's anthropological reconstruction, for they suggest a reading of B 136 that allows for the explanation of the primacy of pure apperception, which Kant asserts here, as the expression of the epistemic function of the latter. This is not to say that the kind of self-consciousness provided by pure apperception has not in fact an enormous anthropological relevance for Kant.¹⁴ Nonetheless, it seems more appropriate to assume that the primacy claim which Kant is making in the transcendental deduction has to do with the role of pure apperception in constituting objective knowledge.¹⁵

To conclude, we can say that there is no need to ascribe to Kant the psychological view that any instance of human self-awareness is dependent on our being capable of pure apperception. This is an anthropological tenet one may attribute to him, but it is by no means necessary to make sense of Kant's claim that in the constitution of objective knowledge, pure apperception has some primacy over inner sense. If, on the other hand, we accept that this primacy claim is related to Kant's epistemological concerns, then we may look for an

¹² This interpretation can be affirmed by the way Kant re-interprets the rationalist terminology of clear and distinct ideas in §§ 5 & 6 of his *Anthropology*.

¹³ In reconstructing this program, we should abstain also from explicating the activity of pure apperception in terms of adopting propositional attitudes.

¹⁴ For this, see §1 of the *Anthropology*, where Kant draws attention to the radical change in a subject's consciousness that occurs the moment a subject starts using the word 'I'.

¹⁵ It is not surprising, against this background, that the more Kant is interested in man's abilities to represent his own mind, the more he focuses on inner sense and its variations. Cf. in particular the *Anthropology* and the late manuscript *On Inner Sense*.

alternative way to make use of Kant's views on self-reflection, which attributes more importance to the passivity involved in man's becoming aware of his mental properties.

4 The possibility of reflective experience

Let me now recall the final question of the second set of interpretive issues specified in section one. Is the primacy of pure apperception, I asked, the only relation Kant assumes between the two kinds of self-consciousness? Given the varieties of self-representations exposed in section two and given, in particular, the option that we can be affected by the play of our thoughts and become aware of this in a passive way, the answer must be 'no'. There must be a sense in which an instance of our thought can be felt or perceived by inner sense, even if it originated from an act of pure apperception. Kant must, in other words, allow that there is a passive way to become aware of one's thought, which in turn amounts to a more robust kind of self-knowledge.

The question might arise whether this is consistent with Kant's attributing epistemic primacy to pure apperception over inner sense. How can we make sense of the notion that we come to know our own thoughts in a passive way without undermining this primacy claim?

Kant does not address this issue at length, but we can get an idea of how he would have addressed it if we take a closer look at his famous reflection, R 5655, entitled *Beantwortung der Frage, ob es eine Erfahrung ist, dass wir denken* (AA 18, 318-20). Is it, the question he is discussing here asks, an experience that we think? Kant begins by clarifying the usage of his notion of 'experience', which he relates to his concepts of 'perception' and of 'aesthetic comprehension'. We do not have an experience, he says, when we are conscious of some representation, but only a perception. Yet, if we come to grasp or comprehend some manifold perceived in one unified representation, then we do have an empirical cognition of an object, and our judgment expressing this cognition does in fact correspond to an experience.

Kant thus defines experience as the judgment affirming an empirical cognition, which is in turn constituted by the unification of some representation provided by the senses. Having established this definition, he goes on, in the second paragraph, to reject the idea that mere consciousness of one's thoughts could possibly constitute experience. Clearly, given the aforementioned conception of experience, neither the imagination of an object nor the having of a cer-

tain thought amounts to an experience. This can be generalized: imagination and consciousness just aren't forms of experience for Kant.

So far, this text consists in a mere elucidation of Kant's concept of experience. In the third paragraph, however, Kant suggests a rather unusual application of this concept, which has some striking implications. After reaffirming that to have an experience requires that one receives some representation from the senses, he claims that even a thought lacking perceptual content may result in an experience if some further and, I think, reflective step is added. Considered as affections or 'determinations' of the mind ('Gemüth' in German), thoughts can be perceived as temporal events which I undergo or as temporal states to which my mind is subjected. Even my thoughts are therefore observable in inner sense and may constitute the content, or in Kant's terminology, the material, of an empirical cognition.

Note that this view implicitly relies on the options pointed out in section two: that, first of all, we can be affected by the products of our capacity for thought; and, secondly, that we can perceive this in a passive way. However, in elaborating this idea of reflexive experience in R 5655, Kant goes even further. He claims that *any* thought we have, whether its content is empirical or merely imaginative, may, when considered as a mental event and observed in inner sense, bring about an empirical cognition which, in turn, when approved by some judgment, results in an experience. Kant's answer to the question whether it is an experience that we think is thus this: Although the concept of experience as established in his epistemology is not straight away applicable to our consciously held thoughts, it is nonetheless possible for us to make an experience on the basis of a thought we have, namely, if we, through a process of reflection, come to see and acknowledge that the thought in question is indeed one we have been entertaining.¹⁶

16 Kitcher (2011, 173-175) has developed an alternative interpretation of this Reflection which denies that the possibility of reflecting on mental events in inner sense is an option for Kant. Her interpretation relies on a particular reading of the second half of the following sentence: "The thought [thinking] itself, although it occurs in time, takes no regard of time when the properties of a figure are to be thought." According to Kitcher, Kant is thinking here of the necessity of an enduring consciousness throughout the combined representation in thought. Whether this is really necessary has been challenged recently by Rosefeldt (2014). My point is different: it seems to me that Kant is thinking of the atemporal properties of the *objects* or *contents* of thought, rather than of some putative atemporal character of the process of thought. Furthermore, I have some difficulties with understanding Kitcher's assumption that processes of conscious combining cannot be given any temporal determination. To be sure, given her account, we have to keep in mind the first representation to be combined during the whole process of combining. We thus cannot say that the first representation occurs at a par-

I would now like to return to the question raised above: How can we make sense of Kant's assumption that we may sometimes become aware of our own thoughts in a passive way without undermining the claim that pure apperception has some kind of epistemic primacy?

Keeping in mind what Kant says in R 5655 about the possibility of making an experience by means of reflecting on the affections of our mind, we are now in the position to suggest a preliminary answer to this. We have to recall, first of all, that inner sense is characterized also as that form of intuition by which we perceive appearances as temporal items (cf. B 37 and B 49f.). Kant himself rehearses this view in R 5655 when he says that any experience, even if it is to represent some atemporal truth, is necessarily correlated with some temporal determination. In having an experience, he says, "I am passive and feel myself affected according to the formal condition of the inner sense" (AA 18, 319).

We can conclude that it seems possible on the Kantian theory to look at every thought or judgment from two sides and to adopt either an atemporal or a temporal perspective on it. To illustrate this, let us assume that I affirm the visual appearance of some lines as configuring a clear and distinct representation of a triangle. I am aware of this appearance and approve it, in an act of pure apperception. I thereby affirm that it is representing a triangle, disregarding the fact that I perceive the lines at a particular time. But I can of course take a step back and look at the mental state I am having *when* approving this appearance as representing a triangle. I will then observe this act of pure apperception as a particular mental episode, and this involves being aware of it as affecting my mind at a certain time. Kant thus seems to adopt some kind of aspect dualism according to which every instance of thought can either be regarded and affirmed as holding independently of any temporal restriction, or perceived as a temporal event.¹⁷

ticular moment of the process of combining. Yet, this is not to say that the whole process cannot be assigned a determinate time, but only that, internally, it has to be structured holistically, rather than as sequence of temporal events. However, I agree with Kitcher about one point: It is hard to see how Kant's notion of inner sense can account for the phenomena she has in mind. This indicates that Kant's notion of inner sense is theoretically under-developed, but not, as she sometimes seems to suggest, that it has no role to play.

17 The question might arise whether it is not crucial that one and the same subject can adopt both perspectives at the same time. Is it not necessary for Kant's views on the possibility of knowledge that we can, at the same time, both perform an act of pure apperception and perceive it as now occurring? This is how I read the emphasis of Henrich 1988, Horstmann 2013 and Emundts 2007 and 2013 that pure apperception is typically some kind of occurrent consciousness. I do not take a position on this question now. What seems more important to me here is to emphasize that these two perspectives necessarily complement each other. They

Now, to safeguard the epistemic primacy of pure apperception over inner sense, these two perspectives cannot be on the same footing; otherwise they would mutually neutralize each other. They must be attributed different theoretical functions.¹⁸ This is indeed what we find in Kant. The two perspectives express, as it were, different concerns of Kant's philosophy. That we can take an atemporal perspective towards all our representations and approve them by an act of pure apperception is a necessary condition of there being mental states that count as objective knowledge. The first perspective is thus absolutely crucial for Kant's transcendental philosophy. The case is quite different if, on the other hand, we want to examine how it is possible for human subjects to improve their outlook on the world by means of philosophical reflection. To address this question, we must be able to consider our thoughts as mental events which are affecting us or have been affecting us. This requires that we adopt the second perspective.

That we can know of our thoughts in a passive way may thus seem unnecessary for Kant's primary epistemological project, which is to exhibit the conditions of the possibility of objective knowledge. But it is, I contend, necessary for another project of his, namely to show that human minds are able to overcome a certain kind of error by means of reflective self-scrutiny. That we can perceive a certain thought as a mental state we have been maintaining so far is necessarily connected with the idea that enlightenment is a possibility for human beings.

5 Self-correction and the requirement for a philosophical conception of self-knowledge

I have shown in this paper that, in contrast to what self-constitutionalist interpretations allege, there is, according to Kant, a way in which we passively become aware of our thoughts. Furthermore, I have argued that, while this possibility may be neglected when Kant's views on knowledge are at stake, it is a crucial epistemic prerequisite of his conviction that human beings are capable for enlightenment.

exhibit different aspects of our mental life, and it is not possible that one may replace the other. They cannot possibly be adopted with regard to the very same aspect of the same item. Whether it is possible for a subject to adopt them at one and the same time is secondary.

¹⁸ This insight goes back to a discussion I had with Richard Moran and Matthew Boyle on a previous version of this text. I am very much indebted to them.

The question might arise why this is relevant for a theory of self-knowledge. Clearly, it is not immediately relevant for the issue that is often discussed in contemporary philosophical debates about self-knowledge: first person authority. The kind of reflective experience discussed in the preceding pages does involve first-person authority, but its analysis does nothing to clarify how, according to Kant, first-person authority is to be understood. For this question, it would suffice to look at conceptions of pure apperception and of inner sense.

However, when we explore the history of philosophy, we see that discussion of self-knowledge has never been restricted to issue of first person authority. Instead, philosophers have often been concerned with self-knowledge for reasons similar to those that motivate Kant's interest in the idea of reflective experience: accounting for the possibility of enlightenment. In more general terms, self-knowledge matters because it is related to the capacity of human beings to recognize their mistakes and to correct their epistemic or moral attitudes. Much more can be said about this on both phenomenological and epistemological levels. In this paper, I have just argued that to keep enlightenment a real option for humanity presupposes that we can become aware of our thoughts in a passive way.

The problem with self-constitutionalism is that this option is neglected, or even undermined, notwithstanding the declared interest of self-constitutionalists in epistemic self-determination. To illustrate why self-constitutionalism cannot do the job, let us have a look at the following example. In my daily life as a teacher of philosophy, it often happens that I am confronted with the fact that I have had mistaken views about a student. Let us assume that I realize that I have been wrong in my judgment that a certain student is disinterested in logic. The crucial question is: What are the preconditions that allow for such an insight?

Without explicating this in detail, we can state that such an insight involves at least two cognitive acts. On the one hand, I have to identify the notion that the student is disinterested in logic as a thought I have been entertaining so far. On the other hand, I have to deny that the student is disinterested in logics. Having distinguished these two acts, it becomes quite clear that the preconditions for my being in the position to acknowledge the wrongness of my former judgment are quite complex. It is not sufficient that I adopt a certain stance towards what is presumably the truth about the student's interest in logic. In addition, I must be able to know what I have in fact have maintained until very recently.

Now, while the possibility of my taking a stance towards some presumed truth may well be accounted for by self-constitutionalism, this is not the case with respect to the other act, the perception of myself as having been in a state

of mind which I now judge as wrong. To acknowledge that I have been wrong, in my view, requires that I can be aware of a thought in a way that does not involve actual endorsement.

We can conclude that either we allow for the kind of reflective experience discussed above and, hence, for the possibility of a passive way of becoming aware of our thoughts, or we have to deny that it may ever happen that we correct our own judgments in the way just sketched. I think this is definitely too high a price to pay, especially for a philosopher; it would undermine nothing less than the very possibility of enlightenment. It is against this background that I am calling for a new kind of approach to the concept of self-knowledge in philosophy, which is not concerned solely with the issue of first person authority, but also helps us understand the possibility of enlightenment, or, in other words, the capacity of humans to recognize their mistakes and correct them by self-reflection.¹⁹

Bibliography

- Allison, Henry 2006 "Kant on freedom of the will", in: Paul Guyer (ed.), *The Cambridge Companion to Kant and Modern Philosophy*, Cambridge: Cambridge University Press, 381-415.
- Bilgrami, Akeel 2006 *Self-Knowledge and Resentment*, Cambridge MA: Harvard University Press.
- Boyle, Matthew 2009 "Two Kinds of Self-Knowledge", *Philosophy and Phenomenological Research* 77, 133-164.
- Brandt, Reinhardt 2013 „Zu Eckart Försters *Die 25 Jahre der Philosophie. Eine systematische Rekonstruktion*“, *Kant-Studien* 104, 367-385.
- Brook, Andrew 2013 "Kant's View of the Mind and Consciousness of Self", *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.),
URL: <http://plato.stanford.edu/archives/fall2013/entries/kant-mind/>.
- Emundts, Dina 2007 „Kant über innere Erfahrung“, in: Udo Kern (ed.) *Was ist und sein soll: Natur und Freiheit bei Immanuel Kant*, Berlin: Akademie-Verlag, 189-205.
- Emundts, Dina (ed.) 2013, *Self, World, and Art: Metaphysical Topics in Kant and Hegel*, Berlin: De Gruyter.
- Emundts, Dina 2013b „Kant über Selbstbewusstsein“, in: Dina Emundts (ed.) 2013, 51-78.

¹⁹ I thank Dick Moran and Matt Boyle for some illuminating discussions on previous versions of this paper. I am also grateful to Dina Emundts, Bernhard Ritter and Yoon Choi for their subtle comments. Bernhard Ritter has also advised me to take a closer look at Reflection 5655. I am particularly thankful to Yoon Choi also for her work on the editing of this paper. Finally, I would like to thank Tobias Rosefeldt for having shared his comments on Pat Kitcher's book as well as the editors for their patience.

- Henrich, Dieter 1988, „Die Identität des Subjekts in der transzendentalen Deduktion“, in: Oberer, Hariolf and Gerhard Seel (eds.), *Kant: Analysen – Probleme – Kritik*, Würzburg: Königshausen & Neumann, 39-70.
- Horstmann, Rolf-Peter 1993 „Kants Paralogismen“, *Kant-Studien* 83, 408-425.
- Horstmann, Rolf-Peter 2013 “Propositional Activity in Kant and Hegel”, in: Emundts 2013, 17-40.
- Kant, Immanuel 1900ff. *Kants gesammelte Schriften. Akademieausgabe*, Berlin/Leipzig: De Gruyter.
- Kitcher, Patricia 2011, *Kant's Thinker*, Oxford/New York: Oxford University Press.
- Klemme, Heiner F. 1996 *Kants Philosophie des Subjekts. Systematische und entwicklungsgeschichtliche Untersuchungen zum Verhältnis von Selbstbewusstsein und Selbsterkenntnis*, Hamburg: Felix Meiner Verlag.
- Moran, Richard 2001 *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton: Princeton University Press.
- Rödl, Sebastian 2007 *Self-Consciousness*, Cambridge MA: Harvard University Press.
- Rosefeldt, Tobias 2014 “Commentary on Chapter 15 of Patricia Kitcher's *Kant's Thinker*”, *Kantian Review* 19, 127-133 doi: 10.1017/S1369415413000344.
- Wright, Crispin 1998 “Self-Knowledge: The Wittgensteinian Legacy”, in: Crispin Wright, Barry C. Smith and Cynthia Macdonald (eds.), *Knowing Our Own Minds*, Oxford: Oxford University Press, 13-46.
- Zobrist, Marc 2011 *Subjekt und Subjektivität in Kants theoretischer Philosophie*, Berlin: De Gruyter.

